T. Pire, B. Espinasse, A. Casali, C. Deco, (2011), « Automatic Extraction of Learning Objects Metadata for Recommendation: a comparative study ». InfoEdu 2011, Convergent Technologies, Integration and Independence Conference, Habana, Cuba, February 7-11, 2011.

EXTRACCIÓN AUTOMÁTICA DE METADATOS DE OBJETOS DE APRENDIZAJE PARA LA RECOMENDACIÓN: UN ESTUDIO COMPARATIVO

AUTOMATIC EXTRACTION OF LEARNING OBJECTS METADATA FOR RECOMMENDATION: A COMPARATIVE STUDY

Taihú Pire[1], Bernard Espinase[2], Ana Casali [1,3] y Claudia Deco[1].

[1] Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario, taihup@gmail.com

Av. Pellegrini 250, (2000) Rosario, Argentina.

[2] Laboratoire de Science de l´Information et des Systèmes – LSIS, Marseille, France, bernard.espinasse@lsis.org

[3] Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas - CIFASIS, Rosario, Argentina. acasali@fceia.unr.edu.ar

RESUMEN

En la última década, Internet se utiliza, entre otras cosas, como fuente de información educativa. Para ayudar en el almacenamiento, clasificación y reutilización de recursos educativos aparece el concepto de objetos de aprendizaje con el fin de clasificar el material educativo, para proporcionar unidades modulares de aprendizaje con metadatos, y para mejorar el acceso y la reutilización de los mismos. En este trabajo se analizan, por un lado, la importancia de los metadatos de los objetos de aprendizaje con el fin de obtener una recomendación personalizada. Por otro lado, se explora el estado del arte de las técnicas de extracción automática de metadatos, y se analizan y comparan diferentes sistemas de extracción. Por último, se presentan algunas conclusiones sobre posibles líneas de investigación para abordar el problema de la falta de información de metadatos en los objetos de aprendizaje.

Palabras Claves: Objetos de Aprendizaje, metadatos educacionales, extracción de información, estándar LOM.

ABSTRACT

In the last decade, Internet is used, among others things, as an educational information source. To help in storage, classification and reuse of educational resources appears the concept of Learning Objects (LO) in order to classify educational material, to provide modular units of learning with metadata, and to improve the access and reuse of them. In this work we analyze, on the one hand, the importance of metadata in Learning Objects in order to obtain a personalized recommendation. On the other hand, exploring the state of the art of automatic metadata extraction, we analyze different software systems and we make a comparison of these systems. Finally, we make some conclusions about several lines of possible research work to address the problem of lack of metadata information in LOs.

KeyWords: Learning Objects, educational metadata, information extraction, LOM standard.

1. INTRODUCTION

Nowadays the web is one of the most important sources of educational material where students and teachers have a big amount of information at their disposal. For this information retrieval process, people use search engines (like Google or Yahoo) which, unfortunately in many cases, do not return the desired information or return a lot of web pages.

To help in storage, classification and reuse of educational resources appears the concept of learning objects. A learning object (LO) is "any digital resource that can be reused to support

learning" [1]. LOs can be used by a student who wants to learn a subject, or may be used by a teacher who wants to prepare materials for his/her class. LOs are described with metadata usually in the standard LOM[1]. Users can retrieve LOs through searches in web repositories. Examples of such repositories are: FLOR[2], Ariadne[3], and OER Commons[4].

Given a topic query, different users have as result the same list of LOs. Generally, he/she checks only the top results, but in many cases these results are not suitable if the search is performed considering only topic keywords. This is because users have different characteristics and preferences, which should also be considered at search time. Recommender systems arise to solve this kind of problem because they can select the material that is most appropriate to user's needs and preferences.

In the following Section 2 we analyze the importance of metadata in learning objects in order to obtain a personalized recommendation. In Section 3, we focus on the automatic LO metadata extraction, analyzing four relevant systems devoted to the LO metadata extraction: SAXEF, TWYS, Looking4LO and MAGIC systems. In Section 4, according to different specific topics, we compare these four automatic extraction systems. Finally, in Section 5 we discuss about this comparative study and we propose some possible research work to address the problem of lack of metadata information in LOs.

2. IMPORTANCE OF LO METADATA

To support users in selecting relevant LOs appropriated to his/her needs and preferences, development of specific recommender systems appears a good way.

In the last years, Artificial Intelligence community has carried out a great deal of work on recommender systems [2]. This kind of systems can help people to find out what they want, especially on the Internet taking into account their personal preferences.

There are some aproaches to achieve the customization of search results taking into account the user profile, including characteristics and preferences ([3], [4]). These systems use LO metadata, with semantic descriptions. Particularly in [5] an educational recommender system is proposed. The architecture presented is based on a multiagent system which allows working with distributed information from LO repositories in a flexible way. The preliminary results obtained are promising in LOs ranking. Nevertheles, the authors claimed that a problem faced is the lack of information in many of educational metadata of the learning objects in the repositories and the metadata quality [6].

In relation to finding appropriate courses in e-learning, Michael Sonntag in [7] analyzes the importance of metadata for learning objects, as these resources may be reused often and possibly in different contexts. Some problems that he points are the lack of metadata, the diversity of standards and the search engine support.

Nevertheless, preparing learning resources with suitable metadata is labor-intensive and consequently there is a lack of quality information in metadata.

An analysis of the information in metadata of LO in some repositories was carried out in [8]. The work considered the repositories: FLOR, OER Commons and Ariadne. Notice that FLOR and Ariadne also support federated search through a federation of repositories. The author points out the lack of information in educational metadata, which is a subcategory of LOM metadata.

Most of the LOs in OER Commons only have metadata specifying the LO Typical Age Range and Language, but neither consider Typical Learning Time (average time needed to follow the LO) nor Interactivity Level. The resources include extra LOM fields such as Type of Contents and Academic Level (similar to LOM metadata: Interactivity Type and Context).

In turn, LOs in FLOR have educational metadata information such as Learning Resource Type, Intended End User Role and Context. Their objects lack of the description of the metadata Interactivity Type and Difficulty, both important to make a personalized recommendation. On the other hand, Ariadne is the most complete repository having LOs with data of almost all educational fields. Regarding the quality of metadata, there are different problems that emerged from the analysis in [8]. For instance, the value of the Language metadata in a LO, was established using only the title, but the body of the LO was in a different Language. Moreover a document was classified as Text in Learning Resource Type metadata but it was a program code and then it should be better classified as Exercise or Example.

[1] http://www.ieee.org
[2] http://www.laclo.org
[3] http://www.ariadne-eu.org
[4] http://www.oercommons.org

Due the importance of metadata for the personalized retrieval of LOs and the lack of quality information of LOs metadata stored in repositories, the development of automatic extraction systems seems to be a very important step towards to solve this problem. In this direction, the goal of this work is to analyze different automatic information extraction methods which are capable of filling some metadata fields and to detect the open research problems in this area.

In the following section, we focus on the automatic extraction of LOs metadata.

3. AUTOMATIC LO METADATA EXTRACTION

Up to now, there are not many works on automatic metadata extraction. Each tool for metadata extraction has its own objectives, architecture and it uses different techniques.

This section presents and analyzes four relevant automatic extraction systems, devoted to the automatic LO metadata extraction: SAXEF, TWYS, Looking4LO, and MAGIC systems.

3.5 SAFEX system

SAXEF (System for Automatic eXtraction of E-learning object Features) ([9], [10]) is a system, created by The Center on Communication Studies (Univ. Palermo, Italy), that automatically extracts the didactic indicators (a sort of DNA) of any web page. In contrast with the other systems, it produces an E-learning Identification Card (EIC) that allows a teacher to easily evaluate whether a page is of interest to him/her.

The SAXEF system has been thought as capable of extracting text/multimedia features from each web page (considered as a learning object) or a group of web pages (which represents a whole course). In practice, given a course or a single learning object, SAXEF produces its EIC. The EICs are organized in a database and are shown through a graphical interface indicating the main topics and their connections.

SAXEF eliminates common words of the text to get relevant words (keywords and topics). For carrying out this task it uses a text file containing articles, prepositions, pronouns, common verbs, etc. Moreover, it identifies relevant words that are inside tags like <title> and <meta> in a HTML file. Then each selected word is provided with a weight. This weight will be used for determining main and secondary topics. In practice, the weight is a score that the word obtains depending on when and where the word appears in the text.

To carry out the multimedia analysis, SAXEF computes the textual and multimedia areas. The textual area is determined by multiplying the number of characters of the web page by the area occupied by each character. The multimedia area is determined by summing up the areas of the multimedia objects present in the web page. In particular, are considered the sizes (in pixels) of images, videos and animations. Moreover, if an audio file is present, its size (in bits) divided by 16 bits (sampling size) is considered. Now, assuming that the area of the web page is the sum of the textual area and the multimedia area, the ratio between the textual area and the total area will provide the analytical index (expressed as a percentage). At the same time, the ratio between the multimedia area and the total area will provide the complementary synthetic index.

To get media types and multimediality level (expressed as a percentage), SAXEF takes into account how many multimedia appear in the web page and the area occupied by them.

3.6 TWYS system

TWYS[5] [11] is a framework developed by Tang Wai Yuen, inside of the Department of Computer Science in the City University of Hong Kong. TWYS is capable of extracting learning object metadata from HTML web pages. It is based strongly in IEEE LOM standard.

TWYS pretends to help users to acquire relevant learning resources easily by enabling search engines to support and adopt IEEE LOM through the automatic extraction from HTML web pages. Some information found on the web page can be mapped to some LOM elements, while others require rules or non-trivial methods to determine the value of the LOM elements. Tang in [11] proposed two methods:

- *Direct Mapping* refers to HTML information that can be mapped with LOM elements directly. Some techniques are StopWord, Term Frequency Weighting (TF-IDF) and Ontology.
- *Heuristic Rules* are applied to other values that cannot be obtained from the information found on HTML web pages. Two techniques were used: (1) Check Existence of HTML Tag, to check whether certain HTML tag exists in a web

[5] We adopt the abbreviation "TWYS" (Tang Way Yuen System) for the framework created by Tang Way Yuen in his Master Thesis.

page or not; and (2) Static Counting of HTML and Content, to count the numbers of certain HTML tags and words in a web page.

First of all, TWYS uses a *Crawler* to collect HTML web pages and then stores them into the local HTML file database. Next, it separates HTML content from the headers and HTML tags through *a HTML Parser*. It also eliminates useless information of the HTML content with Stop Word database, generating a preprocessed file.

Then, the HTML Parser uses the vocabulary predefined from an Ontology (defined to support the classification system) and checks if words of this vocabulary exist in the content of the preprocessed file. If no word of the vocabulary exists in the preprocessed file, it is regarded as not relevant to the domain of interest. The HTML Parser will then discard this HTML web page from the database in order to maintain relevancy of the document repository. If the preprocessed file matches with the domain of interest, the parser will collect all the sub-links in the HTML web pages and feedback to the Crawler to populate the collection. Finally, TWYS extracts and generates LOM records using mapping rules and methods.

TWYS can obtain from HTML Header and Tags the fields Entry, Location, Title, Language, Entity, Date, Format and Size. It uses Stop Words and TF-IDF methods to obtain the features Description and Keyword from the HTML web page content. To produce the fields Purpose, ID, Entry, Description and Keywords, TWYS uses an ontology for the web page content.

As we mention before, TWYS uses some heuristics rules to get the LOM standard educational attributes. To obtain the Interactivity Type, it checks the existence of the HTML tag <Form action=>, <Input>. To retrieve Interactivity Level, it makes a statistic count of the HTML tag <Input>, <action=>. For Semantic Density extraction, it counts the total words and Multimedia of the HTML web page. Finally, to obtain Difficulty field, it counts the distinct words from the web page content.

3.7 Looking4LO system

Looking4LO [12] was created in the Institute of Computation at the Faculty of Engineering (University of the Republic, Uruguay). Looking4LO is a generic and flexible system capable of extracting learning objects with their metadata from XML and HTML files, Word documents, Power Point slides and PDF files, as well as from SCORM Packages.

Looking4LO prototype can extract LOs covering a certain thematic area (for instance: math, logic, history, cooking, etc.) and can automatically provide them with metadata. In the implementation of LOs extraction, Lookin4LO utilizes the platform GATE[6] (General Architecture for Text Engineering). In GATE a document is modeled as its content plus a set of annotations (i.e. additional information about a particular fragment of the document content). Annotations are created and modified by different Processing Resources. In order to generate these annotations, a Processing Resource takes as input the document to process (GATE's internal representation) and might need other resources, for example, an ontology or annotations previously created. The output is the document enriched with new annotations and/or modifications from previous ones. The extraction engine of Looking4LO is integrated by five Processing Resources: *Tokenizer, Sentence Splitter, POS Tagger, Gazetteer and Transducer.*

System inputs are received by a Controller which identifies the type of the documents to be analyzed and then sends it with the rest of parameters entries to the respective processing unit, called Wrapper. Each wrapper handles a different type of document. Inside the wrapper an Extraction Engine is responsible of analyzing the document content. To accomplish this task the extraction process follows the next steps.

First of all, Tokenizer identifies the tokens in the document. Then, Sentence Splitter delimits the sentences. Next, POS Tagger determinates the grammar category of each token. Then, Gazetteer identifies the concepts in the document that belongs to a Domain Ontology. This ontology models the thematic area, this means that the Gazetteer allows detecting which fragments of the document refers to some relevant concept; Transducer executes a set of contextual rules to identify the occurrences of LOs. Finally, the resulting LOs are packaged by the *Packager* in the desired format.

During the retrieving procedure, the metadata extracted for each LO are Author, Read time, Interactivity Level and if the LO has images or not.

3.8 MAGIC system

MAGIC (Metadata Automated Generation for Instructional Content) [13] is a system developed in the IBM Watson Research Center, that automatically identifies segments and generates

[6] GATE (http://gate.ac.uk/) is a framework that allows integrating Natural Language Processing components for the construction of different applications.

critical Metadata according to SCORM standard for instructional content. It can process text and multimedia files.

MAGIC aims to extract SCORM metadata from several file types (e.g., video, audio, text, etc) by applying a suite of particular extracting information methods. Its main target is to assist instructional content authors and course developers with SCORM adoption and enable wider reuse of high value information assets. Here, we are only going to focus over its text metadata extraction method.

MAGIC raises text metadata extraction task with a tool called TEXTRACT [14]. This system is an integral component of a significant infrastructure for document processing and analysis, comprising a number of interconnected, and mutually enabling, linguistic filters. TEXTRACT was designed to perform a variety of linguistic feature extraction functions. Some of these functions are relatively straightforward, e.g. single pass tokenization, lexical look-up and morphological analysis. Others are complex, such as technical terminology extraction, and aggregation of salient phrasal units across large multi-document collections.

MAGIC system follows the steps next detailed. First, text document is tokenized. For this task it using *Frost* (a component of IBM's LanguageWare TM product). Then, each word is tagged with its part of speech (e.g., adjective, noun or verb). Then, analysis modules are applied (here we cannot enter in details because we did not find the suitable information) to extract the following metadata: Title, Keyword, Entity and Description. Finally, a LOM metadata file is generated by populating proper metadata elements in XML format using the extracted information.

3.9 Other systems

In addition to the systems presented, we have also considered other two extraction systems: TextWise and Alchemy API.

TextWise[7] is a platform which allows creating a semantic Digital DNA of documents, revealing the fundamental meaning of text. Semantic Signatures® mines content to uncover the deeper meaning of text and creates a unique "signature" for every document it processes.

Alchemy API[8] is a product of Orchestr8, a company focus on semantic tagging and text mining solutions. It utilizes statistical natural language processing technology and machine learning algorithms to analyze content and to extract semantic metadata: information about people, places, companies, topics, languages among others. API endpoints are provided for performing content analysis on Internet-accessible web pages, posted HTML or text content, and scanned document images.

These two systems, TextWise and Alchemy API, are interesting systems because they are open source and are available to be used to extract some type of metadata. But, as we could not get detailed information about them, in particular what techniques they use, they were not compared with the four extraction systems considered.

4. COMPARATIVE STUDY OF THESE AUTOMATIC EXTRACTION SYSTEMS

This section compares SAXEF, TWYS, Looking4LO and MAGIC systems, according to three main topics: files types treated, metadata extracted, and techniques used for the automatic extraction.

Two main characteristics of the LO metadata extraction can be distinguished: the nature of the files types treated by the extraction and the nature of the metadata extracted.

Files Types: One of the most important characteristics that needs to be considered in metadata extraction systems is that they have to deal with different file types. Example of these file types are: HTML, TXT, PPT, PDF, Word documents and video files.

HTML is a structured text; this means that HTML web page has tags which define its title, language, date and size between others, so it has more information about its content than the others types mentioned before.

Metadata extracted: The metadata extracted by the different systems is one of the most important topic to be considered. Some of the systems presented here extract LOM Standard fields while others extract particular fields defined for a particular approach.

Resources used in the extraction process: For automatic metadata extraction process, each system could use different processing resources, tokenizer, POS tagger, Ontogies, etc.

In the following sub-sections, we compare these four systems, devoted to the automatic extraction of LO metadata, according these three main topics of comparison.

[7] http://textwise.com/
[8] http://www.alchemyapi.com

4.1 Files Types comparison

Most of the extraction systems can manipulate HTML files. In particular, all the six systems presented here can work with HTML web pages. Also, SCORM files are structured and have some metadata fields classified, this type of files is treated by Looking4LO.

Some unstructured files types (e.g. files TXT, PDF, and PPT), meaning that they do not present tags or metadata, are treated by Looking4LO, MAGIC, TextWise and Alchemy API. In particular, if a plain text has a specific structure, that is, it follows certain writing rules or format (for example, if we have lessons composed by a definition, followed by an example and finally an exercise), extractors could take advantage of this situation and makes a better retrieving of metadata.

Others types that are interesting to handle are video and audio file types (e.g. AVI, MPG, MP3, MP4, WMA). MAGIC is the only system presented here that can deal with these formats (others extractors which accomplish this task are Anvil, Elan, EMARaLDA, TASX, MacVisTA. A comparison among them can be found in [14]).

Table 1 shows the file types treated by the different four automatic extraction systems considered. On the one hand, considering the file types all of the systems can manipulate HTML files, because they are easier to handle than other files (e.g. TXT or PPT). There are two systems that can manipulate only web pages (SAXEF and TWYS). In turn, Looking4LO, MAGIC and TextWise can manipulate diverse types of files (e.g. HTML, XHTML, ASP, PHP, PPT). On the other hand, respect to metadata extracted TWYS is the system which extracts the most (in general and educational). Nevertheless, in this approach only the HTML files are treated.

4.2 Metadata extracted comparison

Metadata extracted by the different systems is one of the most important characteristic to be considered. Some of the systems presented here extract LOM Standard fields while others extract particular fields defined for a particular approach. SAXEF does not follow the standard LOM, it produces *an E-learning Identification Card* (EIC) with the following information on the course/object nature: (i) main topics; (ii) secondary topics; (iii) theoretical or practical; (iv) synthetic or analytical; (vi) media types and multimediality level; (vii) complexity level; (viii) links to other EICs with same topics; (ix) links to other EICs with related topics.

Table I: Systems comparison (File Types and Metadata Extracted)

Systems	File types	Metadata extracted
SAXEF	HTML XHTML ASP PHP	Secondary topics Theoretical or Practical Synthetic or Analytical Media types and multimediality level Complexity level Links to other EICs with same topics Links to other EICs with related topics
SWYS	HTML	Entry (LOM 1.1.2) Location (LOM 4.3) Title (LOM 1.2) Language (LOM 1.3) Entity (LOM 2.3.2) Date (LOM 2.3.3) Format (LOM 4.1) Size (LOM 4.2) Description (LOM 1.4) Keyword (LOM 1.5) Purpose (LOM 9.1) ID (LOM 9.2.2.1) Entry (LOM 9.2.2.2) Description (LOM 9.3) Keywords (LOM 9.4) Interactivity Type (LOM 5.1) Interactivity Level (LOM 5.3) Semantic Density (LOM 5.4) Difficulty (LOM 5.8)
Looking4LO	HTML PDF TXT PPT SCORM	Author (LOM 2.3.2) Reading time Has image Interactivity level (LOM 5.3)
MAGIC	HTML PDF	Title (LOM 1.2) Keyword (LOM 1.5) Entity (LOM 2.3.2) Description (LOM 1.4)
TextWise	HTML PDF TXT WORD	Information not available
Alchemy API	HTML TXT	Information not available

Unlike SAXEF, TWYS adopts the LOM standard. It extracts the following fields: Entry (LOM 1.1.2), Location (LOM 4.3), Title (LOM 1.2), Language LOM 1.3), Entity (LOM 2.3.2), Date (LOM 2.3.3) Format (LOM 4.1), Size (LOM 4.2), Description (LOM 1.4), Keyword (LOM 1.5), Purpose (LOM 9.1), ID (LOM 9.2.2.1), Entry (LOM 9.2.2.2), Description (LOM 9.3), Keywords (LOM 9.4), Interactivity Type (LOM 5.1), Interactivity Level (LOM 5.3), Semantic

Density (LOM 5.4) and Difficulty (LOM 5.8). TWYS system is able to extract information respect to the major quantity of fields following LOM standard.

MAGIC only can obtain a subset of metadata extracted by TWYS: Title (LOM 1.2), Keyword (LOM 1.5), Entity (LOM 2.3.2) and Description (LOM 1.4). On the other hand, Looking4LO extracts some LOM fields and other fields of interest: Reading time, Image, Author and Interactivity level. The two lasts metadata belong to LOM standard (LOM 2.3.2 and LOM 5.3).

An interesting point to be stand out, is that some of these systems extract educational metadata and others not. SAXEF's EIC distinguishes when a web page is theoretical or practical and synthetic or analytical. Furthermore, it provides a web page level of multimedia. TWYS produces four educational metadata: Interactivity Type, Interactivity Level, Semantic Density and Difficulty. The extraction of these metadata is not trivial and TWYS uses heuristics rules to retrieve them. In turn, Looking4LO only extracts Interactivity level, and MAGIC does not extract any field.

Table 1 shows the different Metadata extracted by the different automatic extraction systems considered.

4.3 Techniques used comparison

Table 2 shows the different techniques used by the different systems considered to perform their automatic extraction.

Table II: Systems comparison (Processing Resource and NLP tool used)

Systems	Processing resource	NLP tool used
SAXEF	Stop Words Direct Mapping rules Heuristic Mapping rules Statistical Measure	None
TWYS	Ontology Stop Words TF-IDF HTML Parser Direct Mapping rules Heuristic Mapping rules	None
Looking4LO	Ontology Tokenizer Sentences Splitter POS Tagger Gazetteer Transducer	GATE
MAGIC	Tokenizer POS Tagger	TEXTRACT

We precise for each system, the processing resources and the NLP (Natural Language processing) tools used for this extraction.

5. CONCLUSION

In this paper we present the relevance on LO metadata to give a personalized recommendation. The analysis of a variety of repositories has shown that there is a lack of quality information in many of important fields.

Up to now, there are not many works on automatic metadata extraction. Each of the existent tools for metadata extraction has its own objectives, architecture and uses different techniques.

In this report, we try to cover some of these systems, studying and comparing four of them: SAXEF, TWYS, Looking4LO, and MAGIC. We have briefly presented these four extraction systems and then we have compared them according three topics: files types treated, metadata extracted, and resources used for the automatic extraction process.

There are other approaches which staid out of this research report because we could not get detailed information about them. As for example, *OpenCalais*[9] is a rapidly growing toolkit of capabilities that allow to readily incorporate state-of-the-art semantic functionality within blogs, content management systems, websites or applications; and *MetaGlance*[10] is a web service used to generate metadata for web pages, documents and text passages. It automatically adds tags to web pages, quickly gets the gist of what is in a document, and adds metadata to a search index among others things.

We consider that the efforts in automatic extraction of metadata may be increased if we want to have complete and qualified information in the LOs metadata. In particular, the educational metadata extraction offers different difficulties. In this direction there are several open research lines. The systems must be capable to deal with different file types, especially the unstructured ones, and hybrid techniques may be used to deal with the diverse characteristics of metadata.

6. AKNOWLEDGEMENTS

T. Pire acknowledges LSIS for the partial support for his stage in this Laboratory where this work started. A. Casali acknowledges the partial support of project PID-UNR 308.

7. BIBLIOGRAPHY

[9] http://www.opencalais.com/
[10] http://www.metaglance.com/

[1] Wiley, D., "Connecting Learning Objects to Instructional Design Theory: A definition, a metaphor, and a taxonomy", in *D. A. Wiley (ed.) Instructional Use of Learning Objects.* Editorial Association for Instructional Technology, 2002.

[2] Montaner, M., López, B. and de la Rosa, L. 2003. "A Taxonomy of Recommender Agents on the Internet". In *Artificial Intelligence Review* 19: 285-330, Kluwer

[3] Deco C., Bender C., Casali A., Motz R. "Design of a recommender educational system". In Proceedings 3ra. Conferencia Latinoamericana de Objetos de Aprendizaje *LACLO 2008*. México. pp 63-70. Octubre 2008.

[4] Bender, C., Motz, R., Deco, C., Saer, J. "Recuperación personalizada de e-cursos". In Proceedings *IX Congreso Iberoamericano de Informática Educativa*, RIBIE 2008. Caracas, Venezuela, marzo 2008.

[5] Casali, A., Gerling, V., Deco, C., and Bender, C. "A Multiagent System for Personalized Recommendation of Learning Objects". Congreso Iberoamericano de Informática Educativa. IE 2010, Santiago de Chile, Diciembre 2010. In Press.

[6] Casali, A., Gerling, V., Deco, C., and Bender, C., "Sistema Inteligente para la Recomendación de Objetos de Aprendizaje", *LACLO 2010* pp.137-146. San Pablo, Brasil, Setiembre 2010

[7] Sonntag, M., "Metadata in E-Learning Applications: Automatic Extraction and Reuse". In *Christian Hofer, Gerhard Chroust (Eds.):IDIMT-2004. 12th Interdisciplinary Information Management Talks*, pp. 219-231, Universitätsverlag Rudolf Trauner, Linz, Austria, 2004.

[8] Gerling, V., "Un Sistema Inteligente para Asistir la Búsqueda Personalizada de Objetos de Aprendizaje" Degree Thesis on Computer Science, National University of Rosario, Rosario, Argentina, 2009. Available in: www.fceia.unr.edu.ar/lcc/t523/tesina.php?campo1=21

[9] Alfano, M., Lenzitti, B. and Visalli, N., "Text analysis module of a System for Automatic extraction of lEarning object Features (SAXEF)". In *Proceedings of III E-learning Conference*, Coimbra, Portugal, September 2006.

[10] Alfano, M., Lenzitti, B. and Visalli, N., "SAXEF: A System for Automatic eX-traction of learning object Features". *Journal of e-Learning and Knowledge Society*, vol. 3, (2), 83-92, 2007.

[11] Wai Yuen, T., "Automatic Extraction of Learning Object Metadata (LOM) from HTML Web Pages," Master of philosophy, City University of Hong Kong, May 2007.

[12] Hernández, A., Badell, C., Sum, R. and Motz, R., "Convirtiendo el contenido de archivos en objetos de aprendizaje". 2009.

[13] Li, Y., Dorai, C. and Farrell, R., "Creating MAGIC: system for generating learning object metadata for instructional content," in *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia*, pp. 367-370, New York, NY, USA, 2005.

[14] Rohlfing, K., Loehr, D., Duncan, D., Brown, A., Franklin, A., Kimbara, I., Milde, J., Parrill, F., Rose, T., Schmidt, T., Sloetjes, H., Thies, A. and Wellingho, S., "Comparison of multimodal annotation tools workshop report", Gespraechsforschung, vol. 7, 99-123, 2006. Available in: http://www.gespraechsforschung-ozs.de/heft2006/tb-rohlfing.pdf